Int J Life Cycle Assess (2012) 17:689–695
DOI 10.1007/s11367-012-0411-1

UNCERTAINTIES IN LCA

Parameterization in Life Cycle Assessment inventory data: review of current use and the representation of uncertainty

Joyce Smith Cooper · Michael Noon · Ezra Kahn

Received: 23 August 2011 / Accepted: 12 March 2012 / Published online: 3 April 2012

Abstract
Purpose Parameterization refers to the practice of presenting Life Cycle Assessment (LCA) data using raw data and formulas instead of computed numbers in unit process datasets within databases. This paper reviews parameterization methods in the European Reference Life Cycle Data System (ELCD), ecoinvent v3, and the US Department of Agriculture's Digital Commons with the intent of providing a basis for continued methodological and coding advances.
Methods Parameterized data are reviewed and categorized with respect to the type (raw data and formulas) and what is being represented (e.g., consumption and emission rates and factors, physical or thermodynamic properties, process efficiencies, etc.). Parameterization of engineering relationships and uncertainty distributions using Smirnov transforms (a.k.a. inverse transform sampling), and ensuring uncertain individual fractions (e.g., market shares) sum to the total value of interest are presented.
Results Seventeen categories of parameters (raw data and formulas) are identified. Thirteen ELCD unit process datasets use 975 parameters in 12 categories, with 124 as raw data points and 851 as formulas, and emission factors as the most common category of parameter. Five additional parameter categories are identified in the Digital Commons for the presentation and analysis of data with uncertainty information, through 146 parameters, of which 53 represent raw data and 93 are formulas with most being uncertainty parameters, percentages, and consumption parameters.
Conclusions Parameterization is a powerful way to ensure transparency, usability, and transferability of LCI data. Its use is expected to increase in frequency, the categories of parameters used, and the types of computational methods employed.

Keywords Data · Databases · LCA · Parameterization · Uncertainty

Responsible editor: Berlan Rodriguez Perez

Electronic supplementary material The online version of this article (doi:10.1007/s11367-012-0411-1) contains supplementary material, which is available to authorized users.

J. S. Cooper (✉) · M. Noon · E. Kahn
Design for Environment Laboratory, University of Washington,
Box 352700, Seattle, WA 98195-2600, USA
e-mail: cooperjs@u.washington.edu

1 Introduction

Parameterization refers to the practice of presenting Life Cycle Assessment (LCA) data using raw data and formulas instead of computed numbers in unit process datasets within databases. Whereas it has been common practice to use computational models to develop unit process data for some time, recent efforts move computations from supporting documentation to within the datasets themselves.

Consider, for example, the raw data and formulas used by Birkved and Hauschild (2006) to prepare estimates of the fractions of a pesticide emitted to soil, surface water, and ground water for use in a unit process dataset representing crop production. These three model results can simply be reported in unit process datasets within a database, which has been the traditional approach. If instead a parameterized unit process dataset was to be prepared, the raw data (e.g., physical–chemical properties and fate parameters) and formulas (e.g., relationships accounting the primary distribution and the secondary distributions on plants, topsoil, and below the topsoil layer) used by Birkved and Hauschild would be coded in the dataset, providing a wide range of review and analysis capabilities. Using an example provided by Birkved

and Hauschild, the temporal variation of emission fractions can be assessed when the dataset is used in an inventory with those of other unit processes within an inventory. Further, sensitivity analysis can be based on a variation of the model input parameters to reveal their influence on the modeled emission fractions within the context of the overall inventory and impact assessment. Thus, the benefits of data parameterization are transparency (raw data and computations can be clearly documented and reviewed), enhancement of the potential to represent process variants (e.g., variations in load, process efficiency, etc. can be represented), and enhancement of interpretation capabilities (e.g., sensitivity analysis can be performed to the level of internal variables; results can be interpreted as a function of time).

Parameterization has been used within software such as GaBi and SimaPro for some time. It is currently supported by the European Reference Life Cycle Data System[1] (ELCD) through the International Reference Life Cycle Data System (ILCD) data format and will be supported in the upcoming version of the ecoinvent database (v3)[2] through the EcoSpold v2 format. In the EcoSpold v2 data format, parameterized data are included in the flow data category.[3] The description of each parameter and its units are specified in separate XML files that are referenced in the main section via a Universally Unique Identifier (UUID). Parameter names must start with a letter, can only include characters (a–z), numbers (0–9), and underscores (_) and are not case sensitive (i.e., *var1* is interpreted the same as *VAR1*). At a minimum, only the parameter name, the UUID for the parameter identification file, and an amount are required. However, further information such as the parameter units, parameter name, equation, uncertainty information, and comments can be included (see Table S1, Electronic Supplementary Material). Uncertainty information recognizes eight distribution types (beta, binomial, gamma, lognormal, normal, triangular, uniform, and undefined), some descriptive statistics for each distribution, and the unbiased variance of the underlying normal distribution as related to uncertainty based on a pedigree uncertainty (Weidema et al. 2011).

In the ILCD data format, parameterized data falls under the 'Mathematical Relations' category (Joint Research Center 2010). Parameterized data in ILCD have more required fields than EcoSpold v2 (see Table S1, Electronic Supplementary Material), adding the name of the parameter, formula, mean value, and comments/units/defaults as well as a description of the model and use advice (see Table S2, Electronic Supplementary Material). In comparison to EcoSpold v2, ILCD parameter units are specified as comments rather than in a specific units data field; only five uncertainty distributions are recognized (normal, lognormal, triangular, uniform, and undefined); and descriptive statistics are further limited.

Currently, only the ELCD database contains publicly available parameterized inventory datasets (ecoinvent v3 is to be released in 2012). Parameterization relationships in ELCD represent a wide range of engineering relationships for analysis at the inventory level. Datasets being prepared for the LCA Digital Commons[4] also represent a wide range of engineering relationships but also use parameterization to represent uncertainty distributions not currently supported by the ILCD and EcoSpold formats and balance relationships to ensure that fractions sum to totals (e.g., the use of market shares) during uncertainty and sensitivity analyses. The LCA Digital Commons is an open access database and toolset being built by researchers at the University of Washington, sponsored by and in partnership with the United States Department of Agriculture (USDA) National Agricultural Library.

2 Parameterization of engineering relationships in ELCD

In ELCD, 13 parameterized ELCD unit processes[5] represent transportation services (10 datasets), excavators (two datasets), and wastewater treatment (one dataset). Combined, these datasets (see Table S3, Electronic Supplementary Material) include a total of 975 parameters, of which 124 represent raw data (i.e., numbers which should be traceable to documentation of how they were derived) and 851 represent formulas based on raw data and other parameters (i.e., numbers are derived within the unit process dataset).

A review of the 13 datasets allowed 12 categories of parameters to be identified:

1. Concentration Parameters: abundance of a constituent divided by the total volume of a mixture.
2. Consumption Factor Parameters: resource consumption per physical property.
3. Consumption Rate Parameters: resource consumption per time.
4. Consumption Result Parameters: resource consumption per the reference flow.
5. Emission Factor Parameters: emission per physical property or properties.
6. Emission Rate Parameters: emission per time.
7. Emission Result Parameters: emission per the reference flow.
8. Physical or Thermodynamic Property Parameters: a measurable property with a value that describes a physical or

[1] Available at http://lca.jrc.ec.europa.eu/lcainfohub/datasetArea.vm.
[2] Available at http://www.ecoinvent.ch/.
[3] From http://www.ecoinvent.org/ecoinvent-v3/ecospold-v2/.

[4] See http://www.lcacommons.gov/.
[5] See http://lca.jrc.ec.europa.eu/lcainfohub/datasetArea.vm.

thermodynamic state (e.g., mass, volume, density, length, work, and heat).

9. Process Efficiency Parameters: process resource use or conversion per throughput.
10. Process Rate Parameters: process resource use or conversion per time.
11. Percentage Parameters: market share, consumption mix, or production mix.
12. Balance Parameters: ensures a set of market shares, consumption mixes, or production mixes balance to represent 100 % of a total amount.

Parameters in the categories "Consumption Result Parameters" and "Emission Result Parameters" are ultimately unit process exchanges, specifically flows from the technosphere and to nature.

Example parameters in each category and a count of raw data and formulas are presented in Table S4 and Table S5 in the Electronic Supplementary Material. Although this count is certainly a function of the types of processes modeled, the vast majority are emission factors represented by formulas (699 out of the 975 parameters) and used to estimate 82 parameters in the emission results category. Further, whereas parameters in the concentration, consumption rate, process efficiency, and percentage categories are represented only by raw data, parameters in the consumption factor and result, emission result, and balance categories are represented only by formulas. This leaves parameters in the emission rate, physical or thermodynamic property, and process rate categories as mixing the use of raw data and formulas.

The number of parameters in each formula ranges from 1 to 10 with 2 being the most common. Although formulas in the 13 datasets use only basic operators (addition (+), subtraction (−), multiplication (*), division (/), and exponentiation (^)) and *if* statements (i.e., if(b;x;y) returns x if b evaluates to true and y if false), ILCD currently accepts a total of two constants (e and π), 16 operators, and 42 functions (e.g., absolute value, trigonometric functions, logical, and mean, average, minimum, maximum, logarithms, and more).[6] Notably, the vast majority of the formulas used are linear equations, using only addition, subtraction, multiplication, and division.

In the 13 datasets, parameters tend to be given semidescriptive names (e.g., sulphur_ppm is a parameter representing the concentration of sulfur in diesel fuel). Also, because parameter names are sometimes repeated in different unit process datasets (e.g., sulphur_ppm is used in both excavator and the mining truck unit process datasets), it appears that inventory code using these data must separately solve unit process formulas prior to aggregating the data to the inventory.

[6] Based on personal communication with Michael Srocka of Green Delta TC (http://greendeltatc.com/index.html) on June 10, 2011.

3 Parameterization of uncertainty information in the Digital Commons

Given the uses of parameters in the 12 categories as developed for the ELCD data, the Digital Commons also uses parameterization to represent probability distribution functions not currently supported by the ILCD and EcoSpold formats. Here, Commons data representing the production of spring wheat (excluding durum) in Washington State in 2009 are used as an example, with the main data source being the annual USDA Agricultural Resource Management Survey (ARMS[7]). A portion of the example dataset representing the use of synthetic nitrogen fertilizers is provided in the supplemental information (see Table S6 and S7, Electronic Supplementary Material). For this subset of the entire unit process data, there are 146 parameters of which 53 are raw data and 93 are formulas. Also, five additional parameter categories are used:

13. Consumption Parameters: un-normalized resource consumption.
14. Conversion Factors: used to convert a measured quantity to a different unit of measure without changing the relative amount.
15. Emission Parameters: un-normalized emissions.
16. Production Parameters: un-normalized production of product.
17. Uncertainty Parameters: data or formulas needed to represent uncertainty such as relative standard error (RSE), probability, and random numbers for representing uncertainty distributions.

Of the 146 parameters, 27 parameters are categorized as consumption parameters, two as consumption factors, 14 as consumption results, two as conversion factors, one as a physical or thermodynamic property, three as production parameters, 38 as percentages, 12 as balance parameters, and 47 as uncertainty parameters.

3.1 Representing uncertainty distributions

For the very wide variety of engineering relationships potentially useful in preparing parameterized unit processes, there are the commensurately wide variety of supporting raw data and formulas (e.g., variations in feedstock constituents, energy metering data, operating efficiency, and emissions monitoring data). Such data are likely best described by a variety of uncertainty distributions (such as discrete distributions, e.g., Poisson, Bernoulli, etc.; continuous distributions, e.g., normal, lognormal, Weibull, and Chi-square distributions, etc.; and multivariate distributions). Early work by Björklund (2002) and Heijungs and Frischknecht (2005) explains uncertainty distributions as describing how a parameter can be expected to

[7] Data are available at http://www.ers.usda.gov/Data/ARMS/.

deviate from its real value and mention the use of probability and frequency distributions (e.g., normal or Gaussian, and the lognormal distributions), uniform (exact) error intervals, and vague error or triangular intervals. Example methods, case studies, and issues are reviewed by Reap et al. (2008) and Hong et al. (2010). Although more recent LCA research uses well-established statistical methods to evaluate parameters described by probability and frequency distributions (e.g., Birkved and Heijungs 2011; Bojacá and Schrevens 2010; Hong et al. 2010; Ibáñez-Forés et al. 2011; Röös et al. 2010; and Ventura 2011), analyses tend to be modular, such that the results of unit process or sublife cycle models are brought into an inventory as the resulting flow distributions as opposed to automation/simultaneous parameter analysis in inventory and impact assessment. This makes such analyses valuable for the single study in which they are performed, but less so for widespread use as facilitated by databases.

Given a desire to represent uncertainty distributions in datasets within databases, because the LCA data formats (ILCD and EcoSpold) accommodate only beta, binomial, gamma, lognormal, normal, and uniform distributions, Weidema[8] notes that most distributions can be expressed as specific cases of the included distributions, and certainly distributions can be transformed. Within this context and as used in stochastic modeling techniques (see Huijbregts 1998; Lloyd and Ries 2007), a wide range of uncertainty distributions are to be parameterized in the Digital Commons as a function of the uniform and normal distributions that are available in the LCA data formats.

For the first Commons data release, Smirnov transforms (a.k.a. inverse transform sampling) are used to generate random numbers from a continuous probability distribution given its quantile (i.e., its inverse cumulative distribution function) and based on a uniform distribution on [0,1].[9] Because the uniform distribution is supported by the LCA data formats, Smirnov transforms can be used for a very wide variety of distributions within a unit process dataset.

Consider, for example, the ARMS data described by Sommer et al. (1998) as a probability-based survey where each respondent represents a number of farms of similar size and type and the sample data are weighted and expanded to represent operations at the state level. According to Kim et al. (2004), a delete-a-group jackknife is used to estimate the ARMS sample means because the population means are unknown. Differences between a sample and population mean result from nonsampling errors (e.g., related to questionnaire design or data processing) and sampling errors (e.g., related to sample selection, estimation, or nonresponse adjustments). Whereas nonsampling errors cannot be measured directly, a sampling error is represented in ARMS as the RSE of the expected mean and is also called the coefficient of variation.

ARMS RSE data are based on a 15- or 30-sample delete-a-group jackknife, specifically 30 samples for data collected in 2009 and 15 samples for data collected prior to 2009. Because of these relatively small sample sizes, a Student's t distribution is an appropriate representation of the ARMS data probability density functions (Spiegel et al. 2009). Because the Student's t distribution is not supported by the LCA data formats, it is of interest in the Commons to represent the Student's t distributions using parameterization. Given this, the parameterization of the Smirnov transforms begins with the quantile described by Gleason (2000) as developed by Gaver and Kafadar (1984) is used here:

$$Q_t(p;v) \approx \sqrt{v\exp\left(z_p^2 g(v)\right) - v} \tag{1}$$

where p is the probability, v is the degrees of freedom (which is 14 for a sample size of 15 and 29 for a sample size of 30 for the ARMS data), z_p is the inverse standard normal distribution (N[0,1]), and

$$g(v) = (v - 1.5)\Big/(v-1)^2. \tag{2}$$

Because the inverse standard normal distribution, z_p, is not among those available in the current data formats, z_p is estimated here as a function of the inverse error function (erf^{-1}):

$$z_p \approx \sqrt{2}\cdot erf^{-1}(2p-1). \tag{3}$$

Next, because erf^{-1} is not among the list of those available in the current data formats, $\mathrm{erf}^{-1}(2p-1)$ is estimated as described by Winitzki (2003, 2008):

$$erf^{-1}(x) \approx \mathrm{sgn}(x)\sqrt{\sqrt{\left(\frac{2}{\pi a}+\frac{1n(1-x^2)}{2}\right)^2 - \frac{1n(1-x^2)}{a}} - \left(\frac{2}{\pi a}+\frac{1n(1-x^2)}{2}\right)} \tag{4}$$

where:

$$a = \frac{8(\pi-3)}{3\pi(4-\pi)} \approx 0.140012. \tag{5}$$

Solving Eq. 2 at ν=14 or 29 such that $g(\upsilon)$=0.074 or 0.035, the resulting parameterization is achieved for each

[8] Personal communication, March 16, 2011.

[9] Specifically, for a continuous variable x with a cumulative distribution function of $F(x)$, the random variable $y=F(x)$ has a uniform distribution on [0, 1]. Thus, by passing random numbers on the unit interval through the quantile, a sample of a random variable governed by the cumulative distribution function is obtained.

ARMS variable in a manner similar to the example representing the mass of nitrogen applied in 2009 to Washington spring wheat as presented in Table S6 and S7 rows 65–70 in the Electronic Supplementary Material and using parameter names developed by the ARMS developers. As shown, six parameters are used. The first two parameters (Raw_NITLB and RSE_Raw_NITLB) are raw ARMS data representing the weight of nitrogen applied and its RSE, left in English units exactly as downloaded from the ARMS database to allow anyone using or reviewing the data to see the original values. The next parameter, p_t_NITLB, represents the probability (as in Eqs. 1 and 3) as a uniform distribution on [0,1][10] to be used in the Smirnov transform of z_p. Next, zp_t_NITLB uses Eq. 4 and the result of Eq. 5 to approximate z_p based on p_t_NITLB and is then used in the approximation of the parameter t_NITLB as the Student's t value presented in Eq. 1. Finally, NITLB is estimated from the raw data (the mean and the RSE) in a manner similar to the estimation of a confidence interval at the current t value. Thus in an inventory using these data, NITLB is ultimately represented as a Student's t distribution with a mean of 79.08, a RSE of 6.86 %, and a 95 % confidence interval at 79.08±11.08 (for p_t_NITLB=0.025 and 0.975) and

$$NITLB_{95\%} \in Raw\_NITLB\left(1 \pm t\_NITLB\frac{RSE_{NITLB}}{100}\right). \tag{6}$$

Noting that the above formulation overestimates the MS Excel T.INV function (returns the left-tailed inverse of the Student's t distribution given the probability and degrees of freedom) by only 0.23 % from p=0.01 to 0.99 and 0.37 % for p=0.001 to 0.999 for 14 and 29 degrees of freedom, there are certainly other formulations of the Student's t distribution that are candidates for parameterization. For example, early work was developed by Hill (1970) as Algorithm 396, demonstrating precision of over six significant figures to his Student's t approximation (Algorithm 395) and three or four decimal place check values to work that existed at the time. Although Algorithm 396 is still in use today, the code was found here to be difficult to present within the parametric format of unit process datasets. Since Hill published Algorithm 396, example continuing work has come from Dawson (1975) who reached t values within ±5 % and ±8 % for sample sizes of three and four or more and within the range of probabilities from 0.1 to 0.005 and 0.2 to 0.001 and by Koehler (1983) who improved upon Dawson's work for sample sizes of nine and above. More recently, Shaw (2006) explored relationships for even and odd sample sizes and provides very detailed equations up to sample sizes of 21 using Newton–Raphson iterations and ultimately limited by machine precision. As in the case of Hill's Algorithm 396, Shaw's code was found to be difficult to parameterize.

Given these relationships, note that data with RSE values above $1/t$ will have a lower confidence bound below zero, which is not actually possible (e.g., there would not actually be a negative area to which fertilizer is applied nor a negative amount of nitrogen fertilizer applied). Similarly, if the parameter unit of measure is a percentage with a RSE greater than $(100-m_{\%}) / (tm_{\%})$ where $m_{\%}$ is the mean value of the variable, the upper confidence bound will exceed 100 %. To explicitly account for these situations in the parameterized dataset, minimum and maximum values have been placed on several parameters (Table S6 rows 11, 17, 23, 29, 35, 41, 47, 70, and 132). Note, however, that this may not be compatible with existing data formats and inventory code (e.g., EcoSpold recognizes minimum and maximum values only for uniform and triangular distributions).

It is important to note that the above formulation allows uncertainty propagation to be studied as described by Hong et al. (2010) from a single piece of raw data through impact assessment. Reap et al. (2008) describe a best case scenario where all input uncertainty can be represented by probability distributions, uncertainty can be propagated to outputs using well-established techniques, and decision makers can compare statistical differences or expected environmental performance. Although such propagation can be expected to be computationally intensive, it should not be beyond current capabilities and should be worth the effort.

3.2 Balance relationships under uncertainty

The ELCD data uses parameters in the percentage parameter and balance parameter categories to specify the fraction of specific technologies within a group (i.e., market shares) and to ensure the fractions sum to the whole. As mentioned in the notes below Table S4 in the Electronic Supplementary Material, the phrase “Result must be 1!” is provided in the Comments, units, defaults data field in the ILCD format, interpreted here to mean that inventory code must recognize this requirement.

In the Digital Commons, to further ensure that the fractions sum to the whole and to incorporate consideration of uncertainty in the fractions themselves, two alternative percentage balance approaches are used. In the first type, the percentage raw data are accompanied by a RSE. Successive *if* statements are used to balance the set to ensure the total does not exceed 100 % as each data point is varied over its distribution. In the second type, the percentage raw data are represented by a most likely value (sometimes with an underlying distribution) that is assigned a triangular distribution bounded by zero and 100 % as a subjective description of a population represented by the modal value. Again using successive *if* statements, the full set of percentages is balanced so that as each parameter is

[10] Note that ILCD supports the random() function for the generation of a uniform distribution on [0,1] which could be used instead of explicitly specifying the uniform distribution. However, it is not clear if EcoSpold v2 will also support random() and, either way, it must be ensured that the distribution is consistently applied within each estimation of z_p.

varied over the triangular distribution and the set does not exceed 100 %. In both cases, the balance parameters are then combined with the raw data to represent the final value of interest. Note that success of this formulation is dependent upon a sufficient number of samples considered to ensure individual parameters are not biased.

The first type of balance is demonstrated for the mix of nitrogen fertilizer application technologies (with the raw data in Tables S6 and S7 rows 18–47 and the balance equations in rows 48–52). The total set of application technologies are no broadcast and broadcast with and without incorporation as all nitrogen fertilizer or mixed formulations. The second type of balance is demonstrated for the type of synthetic nitrogen fertilizer applied (with the raw data in Tables S6 and S7 rows 78–109 and the balance equations in rows 110–116) and adding data from the USDA's Economic Research Services' national level data on fertilizer use[11] and the nitrogen context of synthetic nitrogen fertilizer from the Natural Resources Conservation Service's Nitrogen Fertilizer Guide.[12] All are a function of NIT_LB (the total pounds of synthetic nitrogen fertilizer applied as *N* as a Student's *t* distribution), with the total set of synthetic fertilizers being anhydrous ammonia, aqueous ammonia, ammonium nitrate, ammonium sulfate, nitrogen solutions (mixtures of urea and ammonium nitrate in aqueous or ammoniacal solution a.k.a. URAN), sodium nitrate, urea, and other nitrogen fertilizers. The nitrogen content of "other nitrogen fertilizers" is represented as those not already accounted for and listed in the Nitrogen Fertilizer Guide (specifically as ammonium thiosulfate, monoammonium phosphate, diammonium phosphate, calcium nitrate, and potassium nitrate) with the most likely value as the average.

4 Discussion and conclusions

From a review of the use of parameterization in ELCD and the Digital Commons, 17 categories of parameters are identified with examples revealing the roles of raw data and formulas. Categories represent the variety of data types used, capturing production, consumption, and emissions relationships; stream constituents; technology use; and conversion factors and physical and thermodynamic properties. The review revealed the following be considered:

- Entering parameter names, raw data, and formulas in the form that they appear in the original data source aids in transparency and data review.
- It is useful to ensure descriptive parameter names (e.g., NITLB instead of *X*).
- Computational instructions in comment fields should be avoided (e.g., the use of the phrase "Result must be 1!" is provided in the ILCD Comments, units, defaults data field) as it is not clear that inventory code will recognize and respond to such instructions, especially if they grow in number and type.
- New uses for the minimum, maximum, and most likely value data fields are identified here, but require inventory code to be developed to recognize them as a part of parameterization.
- Acceptable math notation and functions should be standardized between datasets and inventory software, noting that while simple math notation (+, −, *, and /) is generally consistent between software and programming languages, more complicated functions can differ (e.g., an exponential in excel can be represented as $A \wedge b$ but as pow(A,b) in Java as used in openLCA[13]).
- The acceptability and notation of Boolean logic and conditional statements should be clarified in parameterized dataset documentation to open the possibility of including more complex programming or representative pseudocode, similar to the *if* statements used herein.

Within the context of the parameterization of uncertainty and the current availability of functions, Weidema et al. (2011) note that the choice of distribution has limited influence on the overall uncertainty of a product life cycle due to the aggregation of large numbers of independent variables that will approach a result with a normal distribution according to the Central Limit Theorem. However, we argue that parameterization allows the uncertainty in unit process data to be approximated according to the statistical distributions that are most appropriate to the data. As LCA moves into a phase in which uncertainty is better understood, it is likely that more LCAs will assess uncertainty, and it is, therefore, likely that situations in which the approximation methods matter will arise. Further, it has been our experience that data and models prepared for use in LCA have found uses in other areas of research. This leads us to believe that we have not conceived of the full set of assessments, life cycle and beyond, and that careful preparation of our data can only improve their usefulness.

The use of Smirnov transforms is generally applicable including use with a wide range of standard and nonstandard

[11] These data are available at http://www.ers.usda.gov/Data/FertilizerUse/, and note that geographic specificity is national, thus a larger area than is intended to be represented by the Washington State unit process data, and thus having lower data quality for geographic representativeness.

[12] These data are available in Section 9 of 22 (9e—Nitrogen Fertilizer Guide) at http://www.nm.nrcs.usda.gov/technical/handbooks/iwm/NM_IWM_Field_Manual/Section09/9e-Nitrogen_Fertilizer_Guide.pdf and assuming "nitrogen solutions" can be represented as "mixtures of urea and ammonium nitrate in aqueous or ammoniacal solution" (URAN) as inferred from the Harmonized Tariff Schedule code at http://www.ers.usda.gov/Data/FertilizerTrade/documentation.htm.

[13] See http://www.openlca.org/documentation/index.php/Advanced_functions.

probability distribution functions resulting from error propagated when combining data in small groups or large complex models (e.g., Monte Carlo results). It can be computationally efficient if the cumulative distribution function can be inverted, but may be too computationally expensive for some probability distributions in the unit process standard formats. Within this context the approximations used here provide only a starting place for additional research.

Parameterization holds great promise for the preparation of LCA data, by adding transparency, enhancing the potential to represent process variants, and interpreting study results as well as in the representation of uncertainty. Although herein the use of parameterization is only explored within the context of unit process data, the parameterization of fate and transport data and formulas and impact assessment is also plausible. For example, including raw data representing environmental conditions for fate and transport modeling (temperature, precipitation, wind speed, soil type, and conditions) would be a valuable addition particularly when a large geographic area is being represented. Further, it would be of interest to include raw material abundance data, toxicity data from a variety of essays, etc. in the development of characterization factors.

Current uses of parameterization only begin to take advantage of the true potential. As lists of available constants, operators, and functions become available and are extended, it is anticipated that additional capabilities will be discovered. Of particular interest are methods for representing temporal and geographic specificity within datasets, which are currently not supported in ELCD or the Digital Commons but are expected to be of interest. However, even as capabilities are added, the Commons database and likely others will ultimately include both parameterized and unparameterized unit process data.

Acknowledgments This research was funded by the United States Department of Agriculture National Agricultural Library (agreement number 58-8201-0-149).

References

Birkved M, Hauschild M (2006) PestLCI—a new model for estimation of inventory data for pesticide applications. Ecol Model 198:433–451

Birkved M, Heijungs R (2011) Simplified fate modelling in respect to ecotoxicological and human toxicological characterisation of emissions of chemical compounds. Int J Life Cycle Assess 16(8):739–747

Björklund AE (2002) Survey of approaches to improve reliability in LCA. Int J Life Cycle Assess 7(2):64–72

Bojacá CR, Schrevens E (2010) Parameter uncertainty in LCA: stochastic sampling under correlation. Int J Life Cycle Assess 15(3):238–246

Dawson FH (1975) Alternatives to the use of tabulated values of distributions in statistical programs. Nature 256:148

Gaver DP, Kafadar K (1984) A retrievable recipe for inverse t. Am Stat 38:308–311

Gleason JR (2000) A note on a proposed student t approximation. Comput Stat Data An 34:63–66

Heijungs RR Frischknecht (2005) representing statistical distributions for uncertain parameters in LCA. Int J LCA 10(4):248–254

Hill GW (1970) Algorithm 396 Student's *t* quantiles. Commun ACM 13(10):619–620

Hong J, Shaked S, Rosenbaum RK, Jolliet O (2010) Analytical uncertainty propagation in life cycle inventory and impact assessment: application to an automobile front panel. Int J Life Cycle Asses 15 (5):499–510

Huijbregts MAJ (1998) Application of uncertainty and variability in LCA. Int J Life Cycle Assess 3:273–280

Ibáñez-Forés V, Bovea M, Simó A (2011) Life cycle assessment of ceramic tiles. Environmental and statistical analysis. Int J Life Cycle Assess 16(9):916–928

Joint Research Center (2010) International Reference Life Cycle Data System (ILCD). Documentation of LCA data sets. Ispra, Italy: European Commission. Retrieved from lct.jrc.ec.europa.eu/

Kim CS, Hallahan C, Lindamood W, Schaible G, Payne J (2004) A note on the reliability tests of estimates from ARMS data. Agr Resource Econ Rev 33(2):293–297

Koehler KJ (1983) A simple approximation for the percentiles of the *t* distribution. Technometrics 25(1):103–105

Lloyd SM, Ries R (2007) Characterizing, propagating, and analyzing uncertainty in life-cycle assessment: a survey of quantitative approaches. J Ind Ecol 11:161–179

Reap J, Roman F, Duncan S, Bras B (2008) A survey of unresolved problems in life cycle assessment. Int J Life Cycle Assess 13 (5):374–388

Röös E, Sundberg C, Hansson P (2010) Uncertainties in the carbon footprint of food products: a case study on table potatoes. Int J Life Cycle Assess 15(5):478–488

Shaw WT (2006) Sampling Student's *T* distribution—use of the inverse cumulative distribution function. J Comput Finance 9(4):37–73

Sommer JE, Hoppe RA, Green RC, Korb PJ (1998) Structural and financial characteristics of US farms, 1995: 20th Annual Family Farm Report to Congress. Retrieved from http://www.ers.usda.gov/publications/aib746/

Spiegel MR, Schiller JJ, Srinivasan RA, Alu R (2009) Schaum's outlines—probability and statistics, 3rd edn. McGraw-Hill, New York, NY

Ventura A (2011) Classification of chemicals into emission-based impact categories: a first approach for equiprobable and site-specific conceptual frames. Int J Life Cycle Assess 16(2):148–158

Weidema BP, Bauer C, Hischier R, Mutel C, Nemecek T, Vadenbo CO, Wernet G (2011) Overview and methodology: data quality guideline for the ecoinvent database version 3 (final draft_revision 1). Retrieved from http://www.ecoinvent.org/fileadmin/documents/en/ecoinvent_v3_elements/01_DataQualityGuideline_FinalDraft_rev1.pdf

Winitzki S (2003) Uniform approximations for transcendental functions. Proceedings of the ICCSA—2003, LNCS 2667 (p. 962). Presented at the International Conference on Computational Science and Its Applications—2003

Winitzki S (2008) A handy approximation for the error function and its inverse. Retrieved from http://homepages.physik.uni-muenchen.de/~Winitzki/erf-approx.pdf